Talanta 83 (2011) 1289–1294

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Application of parallel computing to speed up chemometrics for GC × GC–TOFMS based metabolic fingerprinting

Thomas Gröger[a,b], Ralf Zimmermann[b,a,*]

[a] Institute of Ecological Chemistry, Cooperation Group "Analysis of complex molecular systems", Helmholtz Zentrum München, 85764 Neuherberg, Germany
[b] Chair for Analytical Chemistry/Mass Spectrometry Centre – Institute for Chemistry, University of Rostock, 18051 Rostock, Germany

ARTICLE INFO

Article history:
Available online 16 September 2010

Keywords:
Parallel computing
Chemometrics
Comprehensive two-dimensional gas chromatography
Metabolic fingerprinting

ABSTRACT

Parallel computing was tested regarding its ability to speed up chemometric operations for data analysis. A set of metabolic samples from a second hand smoke (SHS) experiment was analyzed with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC × GC–TOFMS). Data was further preprocessed and analyzed. The preprocessing step comprises background correction, smoothing and alignment of the chromatographic signal. Data analysis was performed by applying *t*-test and partial least squares projection to latent structures discriminant analysis (PLS-DA). The optimization of the algorithm for parallel computing led to a substantial increase in performance. Metabolic fingerprinting showed a discrimination of the samples and indicates a metabolic effect of SHS.

1. Introduction

The aim of metabolomics is a comprehensive quantitative and qualitative characterization of the metabolome of a biological system and its dynamics [1,2]. However, due to the large qualitative and quantitative diversity not all components and processes of the metabolome can be analyzed at the same time on one analytical platform. Therefore, different strategies have been established focusing on different biological tasks. Metabolic fingerprinting is focused on a relative comparison of biological systems based on their metabolomic patterns which could be addressed by one experiment or one analytical platform without optimizing the system for a certain small subset of metabolites. The strength of metabolic fingerprinting is its ability to screen and classify huge numbers of samples in short progression. Very common are hyphenated techniques like gas chromatographic–mass spectrometric (GC–MS) or liquid chromatographic-mass spectrometric (LC–MS) couplings [3–6]. The aim of such hyphenation is the separation of different metabolites and matrix before they enter the MS. Considering the complexity of metabolic samples much effort has been made to further increase the separation power of the analytical platforms or to adapt them for a special purpose. Basically, these approaches could be divided into two efforts.

The first one focuses on the analytical platform itself and tries to further enhance the selectivity or separation power of the hardware. With regard to the enhancement of the chromatographic side, higher dimensional separation techniques, like comprehensive two-dimensional gas chromatography (GC × GC) [7–9] in combination with a fast time-of-flight MS, have become very popular over the last years. Due to the introduction of a second orthogonal separation direction, the metabolites become separated over a plane. The increased selectivity of such systems leads to a higher separation power and offers also additional opportunities for data analysis of metabolomic data [10–15].

The second attempt concentrates on the application of chemometrics to further improve the physical/chemical separation. Chemometrics can be applied during data acquisition, data processing and/or data analysis [16–22]. Former attempts for application of chemometrics to GC or MS during data acquisition were utilized e.g. by Phillips [23]. Nowadays, the application of chemometrics for the preprocessing of chromatographic and/or mass spectrometric data [24–27] is more common. The main objectives are the enhancement of the analytical signal and its isolation from interfering signals [12,24,28]. A further field of chemometrics in GC/MS based metabolomics is the statistical analysis of the data [29–31]. The principle objects here are the classification of the different samples according to their metabolite pattern, (semi-)quantification and the identification of discriminating metabolites [32]. In any case, chemometrics has to be applied with care, since complex

* Corresponding author at: Institute of Ecological Chemistry, Cooperation Group "Analysis of complex molecular systems", Helmholtz Zentrum München, D-85764 Neuherberg, Germany. Tel.: +49 089 3187 4544; fax: +49 089 3187 3510.
E-mail address: ralf.zimmermann@helmholtz-muenchen.de (R. Zimmermann).

0039-9140/$ – see front matter © 2010 Published by Elsevier B.V.
doi:10.1016/j.talanta.2010.09.015

issues require a careful selection and interpretation of chemometric tools [33,34].

Both attempts can only be realized at the expense of data size and computationally intensive processing. While higher dimensional separation in combination with fast MS systems produce very large data sets as a consequence of high sample throughput and fast repetition rates of MS detection, chemometrical operations on these data sets can become very intense in resource demands and time, if the complete data set of many samples should be considered.

At the moment only a few vendors offer commercial software for GC × GC (–MS). These packages (e.g. Pegasus, LECO Corporation or GC-Image, Zoex Corporation) provide basic processing or analyzing tools and are highly suitable for target analysis due to a user-friendly and sophisticated graphical user interface. Yet state-of-the-art chemometric operations like proper alignment or multivariate statistics for a comprehensive non-targeted analysis are lacking. In addition, this software does not support state-of-the-art architectures like 64-bit, multi-core processing or emerging techniques like general purpose graphic processing units (GPGPU's). One opportunity could be the application of software packages developed for closely related data sets like two-dimensional gel electrophoresis. Since this sector has a larger sales volume the software is in most cases further developed and it would meet the requirements for the processing GC × GC data files. Recently published work [35] looks very promising but currently the adaptation is not ready for end user application. Appropriate algorithms can also be programmed based on popular programming languages like MatLab, R, and others.

This paper will focus on the implementation of parallel computing [36,37] for analysis of GC × GC–TOFMS data from metabolic fingerprinting to speed up chemometric operations [22,38] based on MatLab.

The main purpose of parallel computing is the ability to either distribute one large data block to smaller blocks or speed up a computer algorithm by distributing different data sets (e.g. from GC × GC–TOFMS) on different workers. Nowadays, the first approach is only relevant for 32-bit Windows systems in which a single application can address only about 3 GB. While data is often collected and stored in lower precision like integer, data processing is often based on double precision operations which increase the space needed in memory size of the data dramatically. Data sets from GC × GC–TOFMS often reach this boarder, at least if multivariate operations are part of the processing. With the introduction of 64-bit architecture and the adaptation of the software, the 3GB boarder has vanished. Now the limitation is the physically available memory of the computer system.

Of much more interest is the ability to speed up data processing by distributing the processing of data to different workers. A requirement for parallel computing is the feasibility to distribute the original data set and to do all further operations on such a distributed set. While the first necessity depends on the used software the second one depends on the structure of the data and the kind of data operation. A problem could be an algorithm which has to access data from the memory of another worker. Such inter worker processes would slow down the overall process due to excessive data transfer. Therefore, as a rule of thumb, the data has to be distributed in such a manner, that all workers can operate on their own. For that reason it could be applicable to redistribute the data set during operation to meet the requirements of each programming step. An example is the alignment of different sample chromatograms to a target chromatogram. Popular algorithms are based on piecewise shifting a small section of a sample chromatogram along a target chromatogram within predefined limits until some quality criterion is optimized. In case of GC × GC such an alignment has to be done in two dimensions. If the chromatogram is distributed among multiple workers it could be necessary to shift a part of the chromatogram from one worker to another, which would break the mentioned rule. For this example a total chromatogram has to be stored on one worker. Still, the whole processing can benefit from distribution, if different workers processed different chromatograms. While such a distribution scheme can be suitable for alignment, it can become a problem, if statistics should be applied to the data set. In such a case quantitative data from different chromatograms but the same time index has to be processed from one worker. In that case, the data has to be redistributed prior to statistics.

Technically, parallel computing is based on multi-core technology. Multi-core processors consist of two or more in most cases identical individual processors. These cores are normally placed within one central processing unit (CPU) and share some of the architecture of the hosting chip. Up to date, dual-core CPU's have come up to a standard in personal computers (PC) and quad- or octo-core CPU's are now commercially available. The gain in performance depends mainly on the used software. In order to take advantage from multi-core architecture, the used software has to divide pending work into different threats which can be processed by different cores. A limitating factor is the ability to divide a task into different threats and the transfer time. The maximum achievable speed-up is described by Amdahl's law [39]. MatLab introduced a parallel computing toolbox to take advantage of local multi-core architecture. However, scripts and data structure have to be modified and optimized for the application of parallel computing.

2. Experimental

2.1. Sample material

Prepared sample material was obtained from Fiehn Labs, Genome Center, UC Davis, CA, USA and had already been analyzed there by GC–MS and FT-ICR-MS and subsequent statistical analysis [40].

Male Sprague–Dawley rats were exposed with aged and diluted side stream cigarette smoke at a concentration of 1 mg/m^3 total suspended particulates for 6 h/d for one (group one, 7 individuals) or 21 days (group two, 7 individuals). There was also a control group with 8 and 7 individuals for each group. (The original experiment includes additional groups from 3 and 7 days exposure).

An aliquot of 30 μL rat plasma was transferred into clean microcentrifuge vial and 400 μL of solvent (isopropanol:acetonitrile:water = 3:3:2) were added. The mixture was vortexed for 10 s and then mechanically shacken for 5 min at 4 °C. After centrifugation at 13,000 × *g* for 2.5 min the supernatant was transferred to new centrifuge tubes and taken to dryness under vacuum and centrifugation. Vials were filled with nitrogen and stored at room temperature until derivatization.

Methyl oxime derivatives were produced by dissolving the dry extracts in 50 μL of freshly prepared *O*-methylhydroxylamine·HCl (40 mg/mL in pyridine). Incubation was done at 37 °C for 90 min under continuous shaking. Subsequent trimethyl silylation was achieved by the addition of 50 μL of *N*-methyl-*N*-trimethylsilyltrifluoroacetamide, followed by continuous shaking for 30 min at 60 °C.

The analysis of variance of the original GC–MS data set by Fiehn Labs from plasma and lung samples, showed that several metabolites were significant at the 0.05 level, including palmitoleic, palmitic and arachidic and cis-2-octadecanoic acid.

2.2. GC × GC–TOFMS

GC × GC–TOFMS analysis was performed on a Pegasus III GC × GC–TOFMS instrument (LECO Corporation, St. Joseph, MI, USA) equipped with cryogenic modulation. The separation on the first dimension was performed on 30 m × 250 μm × 0.25 μm SolGel-1ms. The separation on the second dimension was performed on 2 m × 0.1 μm × 0.1 μm BPX50. All columns were obtained from SGE (SGE Analytical Science Pty Ltd., Ringwood, Australia). 1 μL tempered (15 °C) extract was injected splitless to GC × GC–TOFMS at 250 °C injection temperature. Both columns were housed in the same oven and are subjected to the same temperature gradient (70 °C for 2 min followed by a temperature gradient of 5 °C/min to an end temperature of 300 °C hold for 10 min). The modulation period was set to 1 s. Transfer line and ion source were set to 280 and 250 °C, respectively. Masses were collected from 35 to 600 amu with 100 spectra/second.

2.3. Soft- and hardware

Data acquisition was performed on LECO ChromaTOF software 2.01. Unprocessed raw files were exported as netCDF with an average size of 1.1 GB per file. The complete data set was subsequently imported to MatLab R2009b using built in NetCDF-C interface functions. Data processing includes smoothing, background correction and alignment of the chromatographic signal. Feature reduction and selection was performed using *t*-test, partial least squares/ projection to latent structures discriminant analysis [20,21] (PLS-DA).

Programming was optimized for parallel computing using MatLab and Parallel Computing Toolbox 4.2 for MatLab. Data were optimized for alignment applying MatLab ‘mslowess’ function (with a Gaussian Kernel function) and ‘msbackadj’ function (with a shape-preserving piecewise cubic interpolation as regression method for baseline estimation). For the alignment process we used the well-known COW algorithm [41] and it’s MatLab implementation for one-dimensional chromatographic signals [42] and applied it to two-dimensional GC × GC chromatograms.

Programming was performed on a 64-bit Quadcore System (Intel Core 2 Quad Q9550) equipped with 8 GB Ram.

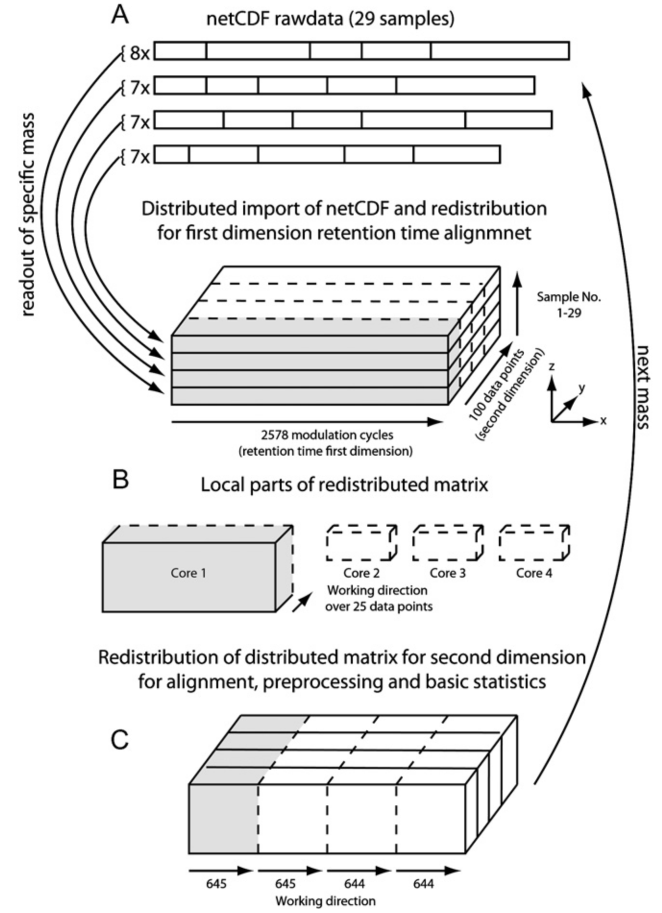
A
netCDF rawdata (29 samples)
8x
7x
7x
7x
readout of specific mass
Distributed import of netCDF and redistribution
for first dimension retention time alignmnet
Sample No.
1-29
100 data points
(second dimension)
z
y
x
2578 modulation cycles
(retention time first dimension)
next mass
B
Local parts of redistributed matrix
Core 1
Core 2
Core 3
Core 4
Working direction
over 25 data points
Redistribution of distributed matrix for second dimension
for alignment, preprocessing and basic statistics
C
645
645
644
644
Working direction

Fig. 1. Distribution scheme optimized for parallel processing. The distribution has to be redistributed to meet the requirements of the actual processing step.

3. Results and discussion

3.1. Parallel programming and speed up

Fig. 1 shows the applied distribution and redistribution scheme for the processing of GC × GC–MS data. We decided to readout and process mass traces successively starting with most comprehensive mass traces like $m/z = 73$. The processing starts with data import and the assembly of an appropriate data structure. First raw data are exported from LECO Pegasus software as netCDF (Network Common Data Form) file format. NetCDF data are very common as system independent exchange format for chromatographic-mass spectrometric hyphenations. The masses and intensity values are stored lineary, so-called variables according to their detection order. While a partial or complete readout is straightforward, the appropriate reshaping of the data can become time-consuming and would benefit from parallel processing. Since the access to the hard drive takes quite a lot of time, any redundant read-write process to the hard drive should be avoided during data processing. The original data set of 29 samples was divided coresponding to their acquisition order into four blocks according to the maximum number of available workers (Fig. 1A). For the import an already distributed three-dimensional dummy matrix was created. The dimensions are chosen according to the nature of the GC × GC chromatogram with first separation dimension as first dimension and second separation dimension as second dimension. Therefore, a single mass trace can be regarded either as a sum of 2578 one-dimensional chromatograms arranged along the first dimension or 100 one-dimensional chromatograms arranged along the second dimension. The mass traces were stored in the third dimension (7× controls from day 1, 8× 1-day exposed rats, 7× control from day 21, 7× exposed rats from day 21). Therefore, every worker has to read out and restructure 7 or 8 raw data sets.

After readout the raw data have to be aligned to a master chromatogram. In case of GC × GC the alignment has to consider a possible shift in two dimensions. The alignment process has to respect the fact, that a disturbance of the chromatographic process could affect different chemical species in a variety of ways. Considering the case that two compounds are only separated by the second dimension, a fluctuation could also affect the separation in the first dimension as well as the separation in the second dimension.

We decided to align the dimensions successively, starting with the first dimension. For that purpose the distributed data matrix had to be redistributed in a way, that any 1D-chromatograms of each 2D matrix of a single mass trace are not split among different workers (*x*-direction) and all chromatograms belonging to the same second dimension time index but to different samples are addressed to the same worker (*z*-direction). Fig. 1B shows an appropriate distribution scheme. The old distribution (solid line) is replaced by a new distribution (dashed line). Each worker has now to align 25 × 28 chromatograms to a target chromatogram.

After the alignment of the first dimension the second dimension also has to be aligned. The distribution scheme had to be redistributed again (Fig. 1C). This processing step also includes a background correction and data smoothing algorithm to prepare data for statistics (*t*-test) which was performed within this distributed processing step as well.

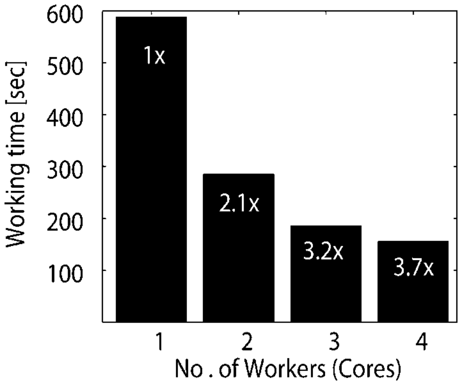

Fig. 2. Gain in performance due to the incorporation of additional workers for one GC × GC mass trace. The absolute gain in performance decreases with additional workers.

Fig. 2 shows the mean speed up due to parallel computing for one mass trace. For this test 10 different mass traces were preprocessed 10 times on 1–4 cores. The given times also comprise the time for saving the results and the processed data matrix to the hard drive. The highest decrease is observed from single-core to dual-core processing. However, this speed up does not only reflect the speed up by parallel computing. If only one core is present, it has to handle also all other processes (e.g. background processes of the operating system) of the host system as well, while an additional core could be exclusively used for data processing. The relative gain in performance decreases with every added core, which is consistent with Amdahl's law.

3.2. *Statistics and metabolic fingerprinting*

Fig. 3 shows the results from data processing and basic statistics. To reduce the number of variables for PLS-DA, the chromatograms from exposed rats were first statistically compared to its control. For this well-defined two class problem a *t*-test was applied as part of the preprocessing to identify significant differences within the chromatograms. Found positions of significant differences ($p < 0.05$) were further input to a PLS-DA analysis in order to test their discriminating strength to separate between the chromatogram of control and exposed rats. The data were preprocessed applying square root scaling and mean centering. The number of latent variables (LV's) were chosen after applying "leave one out" – cross validation. The output of the PLS-DA (loadings and weights) was further used to calculate "variable importance for/influence in projection" – values (vip) [15,20,43]. A reference chromatogram (rat after 21 days exposure) is displayed at the left side (Fig. 3A). To clarify matters, we only show the mass range between 35–250 amu of the GC × GC–MS space. The processing was done for the complete range (35–600 amu). Fig. 3B shows the identified positions, so-called features (vip > 25), within the GC × GC–MS space which are discriminating one day exposed rats from its control group. The same applies to the 21-days group in Fig. 3C. The pattern of the features shown in Fig. 3B and C clearly reflects the fragmentation pattern shown in Fig. 3A. For the sake of a better comparison the fragmentation pattern of a group of peaks is highlighted. Comparing Fig. 3B and C it is also obvious, that the metabolic profile alters with longer exposure period.

To save time, data processing is often done only on selected mass traces. The most prominent would be mass trace $m/z = 73$ because the corresponding ion is most common for TMS-derivatization products. Fig. 4A and B shows the mass trace $m/z = 73$ and the total ion signal (TIC) from the chromatogram shown in Fig. 3A. Even if the TIC chromatogram contains more information the peak pattern would look very similar to the $m/z = 73$ chromatogram. However, if we compare the corresponding feature maps Fig. 4C (only highest vips of $m/z = 73$) with Fig. 4D (highest vips over all mass traces) it becomes obvious that many highly significant features can not be located due to the reduction to prominent mass traces. This indicates the need for comprehensive but also fast data processing.

Fig. 5 reveals clustering of samples based on the PLS-DA scores. If individual mass traces are inspected, time intensive cross validation of the model can be distributed among the workers [22]. For the interpretation of the scores plot it has to be considered, that neither a test for outliers nor a correction of the raw data via internal standards or normalization was performed. We intentionally excluded this kind of quality control, since it is very time-consuming and therefore not always compatible with high throughput analysis. Control and treated individuals could be separated for both cases using LV 1 for day one and LV 1 and 2 for day 21. For both cases the individuals with numbers 1 and 11 are separated from their group. The inspection of Q-Residuals and Hotelling's T2 indicate possible outliers. An additional reason for the poor clustering could be the biological diversity of the individuals which could be associated with a different response to SHS.

3.3. *Outlook for further acceleration of data processing*

Due to the limitations given by Amdahl's law, Fig. 2 illustrates that an additional distribution of data to a limited number of additional workers will only slightly increase the overall process. To gain a noticeable acceleration a huge number of additional workers have to be added. A realization based on CPU's would be very expen-

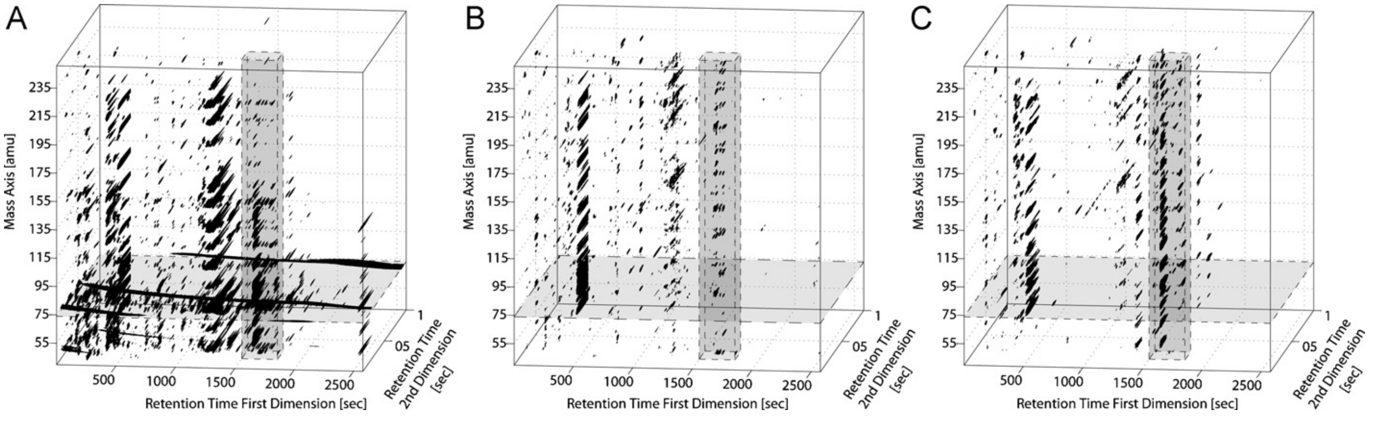

Fig. 3. The left picture shows the chromatogram obtained from the GC × GC–TOFMS analysis of rats' plasma. Figure (B) and (C) indicate discriminating features for 1-day SHS exposure and 21-days SHS exposure. The vertical beam denotes the position of some discriminating features. The horizontal plane indicates $m/z = 73$ (refer Fig. 4).

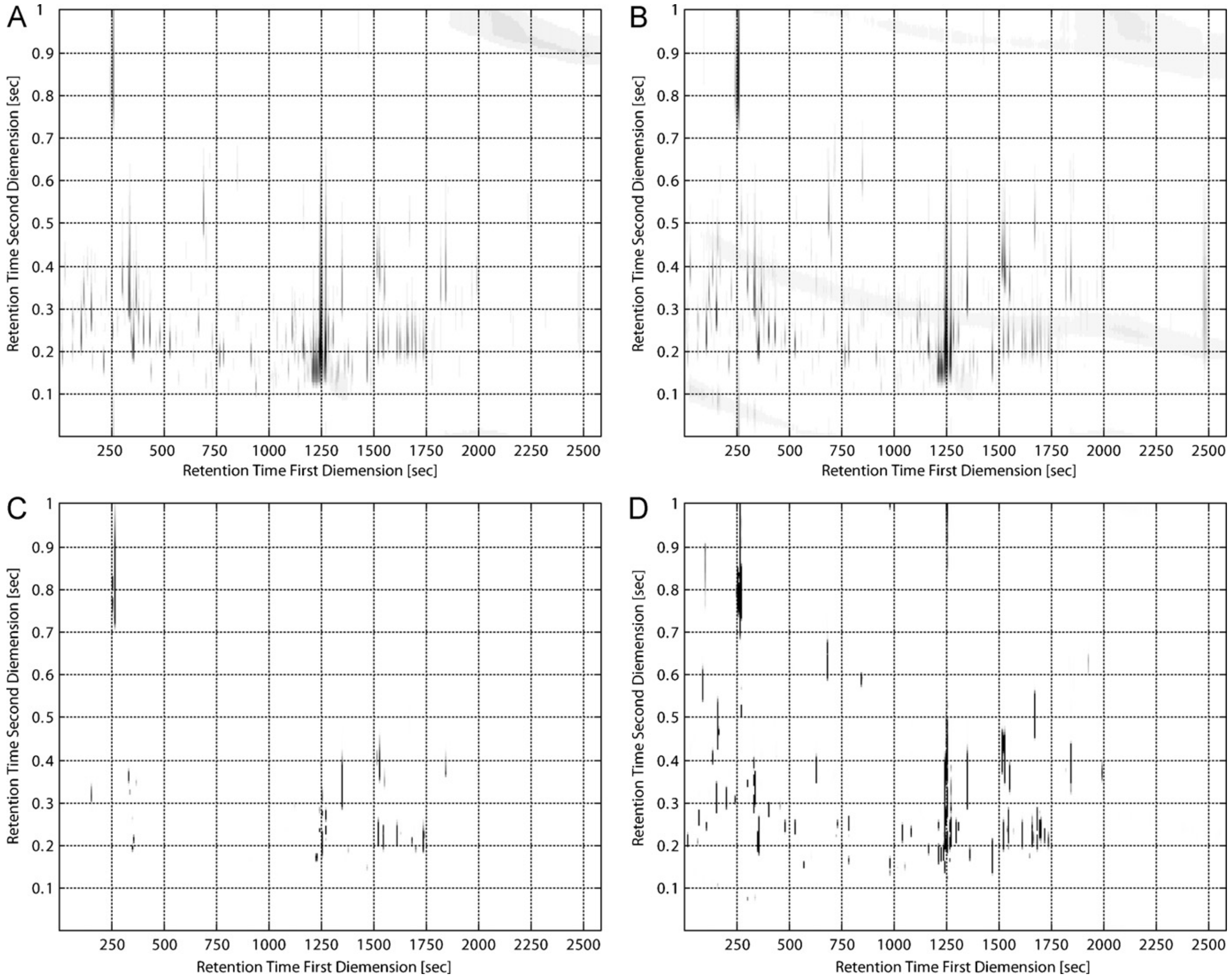

Fig. 4. Figures A and B show the mass trace $m/z = 73$ and the TIC signal from the chromatogram also shown in Fig. 3. Figure C show the detected features found on mass trace $m/z = 73$. Figure D shows the increase in information (more features are found) if the complete data set is analyzed.

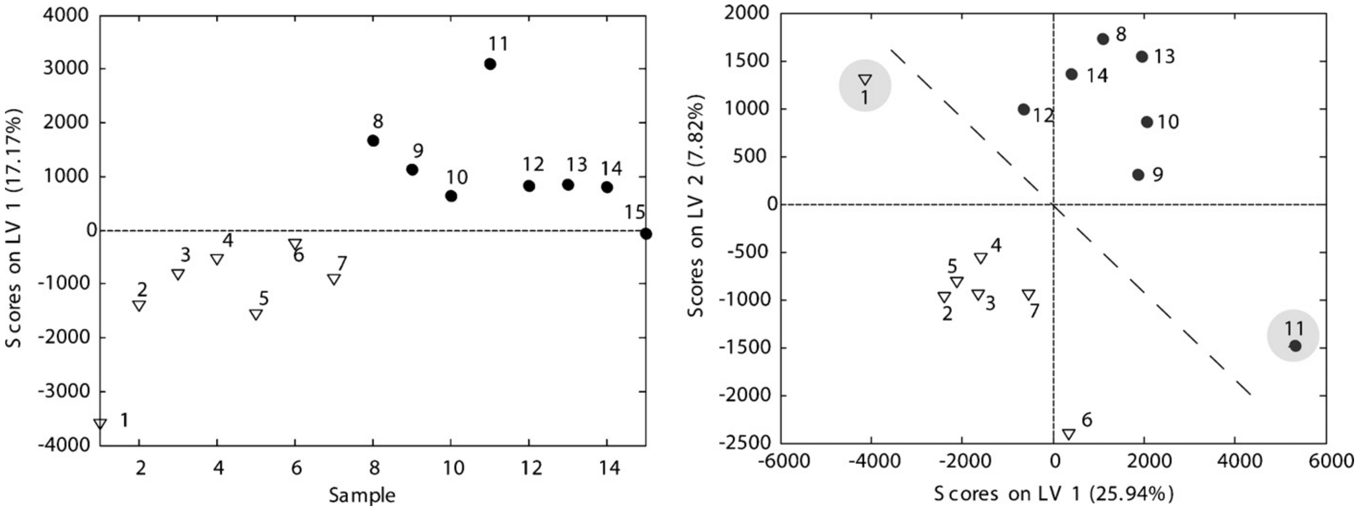

Fig. 5. Scores plot from PLS-DA. Left: day one (triangle: control; circle: SHS treated), right: day 21 (triangle: control; circle: SHS treated).

sive. An affordable opportunity could be the access of graphical process units (GPU's). GPU's are usually located on the graphic card of computer units. Due to their focus on graphical operations they are designed for high performance matrix and vector operations. Unlike CPU's they consist of hundreds of cores which are structured in parallel. By the help of general purpose GPU's (GPGPU's) and the development of the compute unified device architecture (CUDA) by NVIDIA it is now possible to incorporate the massive floating point computationally of these processors for a broad range of applications. CUDA could further be implemented into MatLab via commercial and free software. It is now possible to outsource intense operations from the CPU to GPU and take advantage from GPU resources. Depending on the kind of computer operations a gain in performance up to ×100 is reported in literature. First experiments look very promising, however only basic operations could be addressed so far.

4. Conclusions

Modern analytical platforms like GC × GC–TOFMS in combination with sophisticated data processing and chemometric data analysis require strong computational power to obtain the ability to handle large amounts of data in appropriate time. The application of parallel computing is an attractive and affordable opportunity. Sophisticated programming substantially speeds up the overall process and allows resource-intensive chemometric operations. The time saving enables a more comprehensive investigation of the data. New techniques like GPGPU's will further increase the efficiency of processing and become very attractive considering the cost-performance ratio.

Acknowledgements

Experimental work on rat exposures and sample preparations were funded by a NIH grant to OF from the U.S. National Institutes for Environmental Health, 5R01 ES013932. We thank Marion Hurka, University Konstanz, for copy-editing our manuscript.

References

[1] O. Fiehn, Plant Mol. Biol. 480 (2002) 155–171.
[2] W. Weckwerth, Annu. Rev. Plant Biol. 54 (2003) 669–689.
[3] O. Fiehn, Trends Anal. Chem. 27 (2008) 261–269.
[4] K.K. Pasikanti, P.C. Ho, E.C.Y. Chan, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 871 (2008) 202–211.
[5] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass Spectrom. Rev. 26 (2007) 51–78.
[6] S.G. Villas-Bôas, S. Mas, M. Åkesson, J. Smedsgaard, J. Nielsen, Mass Spectrom. Rev. 24 (2005) 613–646.
[7] J.B. Phillips, J.N. Xu, J. Chromatogr. A 703 (1995) 327–334.
[8] P. Marriott, R. Shellie, Trends Anal. Chem. 21 (2002) 573–583.
[9] J. Beens, U.T. Brinkman, Anal. Bioanal. Chem. 378 (2004) 1939–1943.
[10] R.A. Shellie, W. Welthagen, J. Zrostliková, J. Spranger, M. Ristow, O. Fiehn, R. Zimmermann, J. Chromatogr. A 1086 (2005) 83–90.
[11] W. Welthagen, R.A. Shellie, J. Spranger, M. Ristow, R. Zimmermann, Metabolomics 1 (2005) 65–73.
[12] R.E. Mohler, K.M. Dombek, J.C. Hoggard, K.M. Pierce, E.T. Young, R.E. Synovec, Analyst 132 (2007) 756–767.
[13] K.M. Pierce, J.L. Hope, J.C. Hoggard, R.E. Synovec, Talanta 70 (2006) 797–804.
[14] M.F. Almstetter, I.J. Appel, M.A. Gruber, C. Lottaz, B. Timischl, R. Spang, K. Dettmer, P.J. Oefnert, Anal. Chem. 81 (2009) 5731–5739.
[15] X. Li, Z. Xu, X. Lu, X. Yang, P. Yin, H. Kong, Y. Yu, G. Xu, Anal. Chim. Acta 633 (2009) 257–262.
[16] K.M. Pierce, J.C. Hoggard, R.E. Mohler, R.E. Synovec, J. Chromatogr. A 1184 (2008) 341–352.
[17] B. Lavine, J. Workman, Anal. Chem. 80 (2008) 4519–4531.
[18] K. Lavine Barry, J. Workman (Eds.), Chemometrics and Chemoinformatics, Washington DC, 2005, pp. 1–13.
[19] J. Trygg, E. Holmes, T. Lundstedt, J. Proteome Res. 6 (2006) 469–479.
[20] K.R. Lee, X. Lin, D.C. Park, S. Eslava, Proteomics 3 (2003) 1680–1686.
[21] L. Ståhle, S. Wold, J. Chemometr. 1 (1987) 185–196.
[22] A.d.S. Soares, R.K.H. Galvão, M.C.U. Araújo, S.F.C. Soares, L.A. Pinto, J. Braz. Chem. Soc. 21 (2010) 1626–1634.
[23] J.B. Phillips, Anal. Chem. 52 (1980) 468A–478A.
[24] J.H. Christensen, J. Mortensen, A.B. Hansen, O. Andersen, J. Chromatogr. A 1062 (2005) 113–123.
[25] H.H. Kanani, M.I. Klapa, Metab. Eng. 9 (2007) 39–51.
[26] M. Katajamaa, M. Oresic, J. Chromatogr. A 1158 (2007) 318–328.
[27] J. Boccard, J.-L. Veuthey, S. Rudaz, J. Sep. Sci. 33 (2010) 290–304.
[28] C.G. Fraga, J. Chromatogr. A 1019 (2003) 31–42.
[29] T. Groeger, M. Schaeffer, M. Puetz, B. Ahrens, K. Drew, M. Eschner, R. Zimmermann, J. Chromatogr. A 1200 (2008) 8–16.
[30] P. Jonsson, J. Gullberg, A. Nordstrom, M. Kusano, M. Kowalczyk, M. Sjostrom, T. Moritz, Anal. Chem. 76 (2004) 1738–1745.
[31] M. Bylesjo, M. Rantalainen, O. Cloarec, J.K. Nicholson, E. Holmes, J. Trygg, J. Chemometr. A 20 (2006) 341–351.
[32] R. Madsen, T. Lundstedt, J. Trygg, Anal. Chim. Acta 659 (2010) 23–33.
[33] M. Daszykowski, B. Walczak, TrAC Trends Anal. Chem. 25 (2006) 1081–1096.
[34] C.D. Brown, R.L. Green, TrAC Trends Anal. Chem. 28 (2009) 506–514.
[35] H.-G. Schmarr, J. Bernhardt, J. Chromatogr. A 1217 (2010) 565–574.
[36] O. Trelles, Brief Bioinformatics 2 (2001) 181–194.
[37] G. Vera, R. Jansen, R. Suppi, BMC Bioinformatics 9 (2008) 390.
[38] M. Horoi, R.J. Enbody, Int. J. High Perform. Comput. Appl. 15 (2001) 75–80.
[39] X.H. Sun, Y. Chen, J. Parallel Dist. Comput. 70 (2010) 183–188.
[40] S. Datta, K. Pinkerton, J. Joad, O. Fiehn, Metabolic Alterations from Acute Exposure to Second-hand Smoke (SHS) in Rat Lung and Blood Plasma as Determined by GC–TOF and FT-ICR Mass Spectrometry, San Francisco, CA, 2007.
[41] N.-P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, J. Chromatogr. A 805 (1998) 17–35.
[42] G. Tomasi, F.v.d. Berg, C. Andersson, J. Chemometr. 18 (2004) 231–241.
[43] S. Wold, M. Sjöström, L. Eriksson, Chemometr. Intell. Lab. Syst. 58 (2001) 109–130.